Efficient Machine Learning under Time Constraint

氏名（本籍地）	小林　隼人
学位の種類	博　士（情報科学）
学位記番号	情　博　第　４７５　号
学位授与年月日	平成２２年　３月２５日
学位授与の要件	学位規則第４条第１項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）システム情報科学専攻
学位論文題目	Efficient Machine Learning under Time Constraint (時間拘束条件下における効率的な機械学習)
論文審査委員	（主査）東北大学教授　篠原　歩 東北大学教授　田中　和之　東北大学教授　石黒　章夫（工学研究科）

論文内容の要旨

Machine learning is used to enable computers to achieve *learning*, which is the intellectual ability of humans to acquire new knowledge or skills from their experiences. In the field of *Artificial Intelligence (AI)*, there have been many studies conducted on machine learning, which covers a broad range of applications, such as web search engines, spam detection, finance prediction, pattern recognition, clustering DNA sequences, and autonomous robots.

Considering actual applications in the real world, we must often conduct machine learning under tight time constraints. Robot learning, that is machine learning on robots, is a fine example of problems with tight time constraints, due to physical movements on robots. Adaptive robots, which can change their behavior based on data from environments, are expected to perform well at different situations such as
hazardous tasks, disaster relief, medical services, household chores, as well as industrial fields, and thus the field of robot learning has recently been attracting a great deal of attention.

In this thesis, we investigate efficient machine learning under time constraint from three different viewpoints:

- *Machine* ··· how to improve machine learning algorithms,
- *Environment* ··· how to construct effective experimental environments, and
- *Human* ··· how humans help machines to learn, if possible.

Viewpoint of Machine

This viewpoint is the most common approach of studies on machine learning.

This thesis focuses on machine learning methods that allow robots to acquire some skills based on the interaction with environments and/or humans. Genetic algorithms and reinforcement learning algorithms are representative ones. The former searches the maximum or minimum of an objective function, while the latter approximates a policy function itself, although they are often applied to the same problem. According to Sutton and Barto, reinforcement learning algorithms are efficient for problems with Markov Decision Processes (MDPs), where a robot can perceive and utilize its state, such as path planning. By contrast, genetic algorithms are suitable for instantaneous problems, such as acquisition of ball kicking motions. For achieving case studies in our studies, we will adopt an appropriate machine learning method depending on a target problem.

We consider machine learning as an optimization process, whose objective function has time-consuming procedure, e.g., physical movement of robots. In this case, we utilize meta-heuristics such as genetic algorithms, since we generally cannot know the formula of objective functions. For directly reducing evaluations, i.e., calls of such an objective function, there have been many studies on *surrogate approach* that often evaluates an approximate model instead of an objective function. Commonly-used models are neural network models and Kriging models. There also have been many studies conducted on another approach that indirectly reduce evaluations by noise reduction. Note that those approaches are meta-strategies and can be applied to any sampling algorithm.

Viewpoint of Environment

The experiments on real robots, especially quadruped robots, take much more time and cost than those on PCs. It is also an enormous difference since we often need to consume a lot of energy for treating physical objects such as robots. The use of virtual robots is one of the efficient methods to avoid those difficulties, and there are many studies on dynamic simulated environments. However, since simulated environments cannot produce complete, real environments, we finally need to conduct experiments in the real environments where basic skills heavily depend on complex physical interactions.

In experiments with real robots, *autonomous learning*, by which robots acquire some skills on their own without human intervention. There have been many studies conducted on the autonomous learning of quadrupedal locomotion, which is the most basic skill for every movement. However, studies on autonomous learning of the skills used to control the other objects such as balls have not been studied as much as quadrupedal locomotion. Although these studies are usually case studies, they are regarded as important achievements to verify the performance of learning methods on a real robot.

Viewpoint of Human

Learning from human demonstrations is quite effective to reduce evaluations. Whereas most studies experimentally analyze the improvement of learning by human demonstrations from the viewpoint of machine, in this thesis, we directly focus on teaching strategy of humans. We theoretically analyze the behavior of teaching in a framework of *computational learning theory*, which is used to mathematically formulate a model of learning. In the field of computational learning theory, many studies have been conducted using models such as inductive inference, PAC learning, and query learning. Further, many studies have been carried out on teaching, which is inextricably linked to learning. To our knowledge, there are no study directly conducted on teaching under time constraint in this research area.

Our Contributions

In Chapter 2, for demonstration experiments with real robots, we construct a flexible framework suited for real-time embedded robot system. The framework allows us to easily exchange modules depending on our needs in plug-in fashion and to intuitively describe robot control programs to use the modules with a scripting language.

From the viewpoint of machine, in Chapter 3, we propose a meta-strategy, *thinning-out*, to skip-over unpromising points to be evaluated on an objective function. We investigate properties of thinning-out on various test functions and establish a connection to a competitive meta-strategy, surrogate approach. As an actual application of thinning-out, we address learning of shooting (or ball kicking) motions by virtual robots. Our experimental results show that robots can learn sophisticated shooting motions, which are much different from the initial motion, within a feasible number of evaluations. These results suggest that thinning-out can work well in practical problems.

From the viewpoint of environment, in Chapter 4 and Chapter 5, we consider how to save human time in experiments with real robots. As a case study, we tackle learning of goalie strategy in soccer. We introduce one dimensional model for trapping (or grasping an oncoming ball) skills, and propose a method by which robots can autonomously learn the skills without human intervention. In our experiments, we verify that knowledge sharing among robots can accelerate their learning, even though shared knowledges are rather simple due to their limited computational resources. Further, we develop an augmented environment for learning of two dimensional goalie strategy, which does not include trapping skills. This environment allows real robots to autonomously learn the strategy by using a virtual ball. This study is important as a rare case of robot learning in an augmented environment.

From the viewpoint of human, in Chapter 6, we consider a case where a teacher can help a learner to learn some concept within a time limit. A good instance is learning of goalie strategy by human-robot interaction, where a human rolls a real ball to a real robot. We formulate a theoretical model of such a case, where the complexity of a target concept is measured in terms of the *optimal teaching error*, i.e., the optimal worst-case error. We say that a concept class is *optimally incremental teachable* if the teacher can optimally teach it to the learner whenever teaching is terminated. Further, we define the *teaching dimension with error*, which, in short, is a dual complexity of the optimal teaching error. We give the exact analysis of the optimal teaching error, the optimally incremental teachability, and the teaching dimension with error on natural concept classes such as monomials. Our analysis supports our intuition that it is sometimes necessary to tell a lie in teaching within a time limit.

論 文 審 査 結 果 の 要 旨

機械学習は「学習」という人間の高度な思考形態を計算機上で実現するものであり、人工知能の分野で様々な研究が行われてきた。機械学習の実世界への応用を考えたとき、しばしば厳しい時間拘束条件を考慮する必要がある。例えばロボット動作の学習では、その物理動作に時間がかかるため、実現可能な動作回数は数百回程度に制限されてしまう。著者は、この時間制約条件下において機械学習を効率化する手法について、3 つの視点――機械視点・環境視点・人間視点――から研究を進める。本論文はこれらの成果をとりまとめたもので、全編6章からなる。

第1章は序論である。

第2章では準備として、ロボットを使った実証実験を容易にする、拡張性・柔軟性に優れたプログラミング・フレームを構築する。このフレームワークにより、開発したモジュールを簡単にプラグインとして追加し、スクリプト言語で直感的にロボットプログラムを記述することができる。このフレームワークは3章以降で実際に活用されており、極めて実用的である。

第3章では、機械視点、すなわち機械学習アルゴリズムの改善という観点から、最適化プロセスにおける目的関数の評価回数を削減するためのメタ戦略「間引き」を提案する。多種多様なテスト関数を用いた実験により、評価回数が制限されているとき、間引きによりサンプリング手法の性能が劇的に向上することを示す。さらに間引きの実応用例として、仮想環境上のロボットによるシュート（ボールを蹴る）動作の学習に取り組む。これはメタ戦略研究の新たな分岐を開拓した稀有な研究であり、高く評価できる。

第4章と第5章では、環境視点、すなわち効果的な実験環境の構築という視点から、事例研究としてロボットサッカーにおけるゴールキーパー戦略の学習に取り組み、ロボットが全く人手を介することなく自律的に学習を進めるための 2 つの実験手法を提案する。ひとつは 1 次元のボール捕獲戦略をスロープにより自律的に学習させる手法であり、もうひとつは拡張環境上の仮想ボールを活用することで、ボール捕獲戦略の自律的学習を 2 次元で実現する手法である。これは事例研究として優れており、特に拡張環境を機械学習に活用する研究は独創的である。

第6章では、人間視点、すなわち人間の教示による学習の効率化という観点から、時間拘束条件下における教示の理論モデルを計算学習理論の枠組みで定式化する。モデル上の教示の複雑さを評価するために、最適教示誤差、最適漸増教示可能性、誤差付き教示次元という 3 つの指標を定義し、単項式の概念クラスなどにおけるそれらの指標の厳密な解析を与える。解析結果は、矛盾した例を与える教師が最適になる場合があることを示しており、教示する時間が制限されているとき、小さな例外を無視して概要だけを伝えた方が相手の理解が進みやすいという直感によく一致している。これは、教示の理論で初めて時間拘束条件を考慮した先駆的研究であり、今後の発展にも期待できる。

以上要するに本論文は、時間拘束条件下における機械学習を理論と応用の両面から研究し、機械、環境、人間の 3 つの視点から効率化を実現したものであり、情報科学、特に機械学習の発展に寄与するところが少なくない。

よって、本論文は博士（情報科学）の学位論文として合格と認める。